# Link of published website:

# <https://shdwfruit.github.io/ds4200wp2/>

# Visualization Design Explanation

## Count Distribution of Cut, Color, and Clarity (Interactive Bar Chart)

This interactive bar chart offers an overview of how diamonds are categorized by cut, color, and clarity. By allowing users to toggle between the three variables, the chart helps establish a foundational understanding of the dataset’s composition—such as the dominance of Ideal cuts and color grade G. We placed this visualization at the beginning to orient the audience with categorical distributions before diving into deeper analytical relationships. It sets the stage for later comparisons involving these key attributes.

## Distribution of Numerical Variables: Carat, Depth, and Table (Interactive Histogram)

This visualization displays the frequency distributions of three important numerical variables—carat, depth, and table—through an interactive histogram. Users can switch between the variables to observe individual patterns, such as the right-skew of carat size and the centered distributions of depth and table. Positioned early in the EDA section, this chart complements the earlier categorical bar chart, providing a full picture of how both types of features are distributed and preparing users to interpret their influence on price.

## Sankey Diagram of Cut, Color, and Clarity

The Sankey diagram visualizes how cut, clarity, and color categories flow into one another, with line thickness representing the volume of diamonds in each pathway. This provides a powerful overview of how attribute combinations co-occur in the dataset, such as Ideal cut diamonds being commonly linked to VS2 clarity and G/H color grades. We placed this visualization directly after the simpler distribution charts to smoothly transition into more complex multi-feature relationships while keeping the user's focus on categorical variables.

## Correlation Matrix Heatmap

The correlation heatmap shows the strength and direction of relationships between numeric features, using color gradients to highlight strong positive correlations like that between carat and price. Weak or negligible correlations, such as depth and price, are also clearly visible. We positioned this visualization right before introducing price-focused plots, as it offers statistical context that helps explain why certain features (like carat and physical dimensions) are analyzed more closely in relation to price later on.

## Scatter Plots: Price vs x/y/z Dimensions

These scatter plots visualize how each physical dimension (length x, width y, depth z) of a diamond relates to its price. The trend shows that larger dimensions generally correspond to higher prices, reinforcing the earlier correlation findings. These plots are placed in the “Physical Dimensions” section to give users a direct, visual understanding of how diamond size impacts value. They also help distinguish between the individual roles of x, y, and z—despite all being highly correlated with carat.

## Scatter Plot: Price vs Carat by Cut

This scatter plot presents the relationship between price and carat weight, colored by cut category, to show how different cuts affect pricing at various carat levels. Premium cuts, for example, exhibit higher price trends at larger carats compared to Ideal or Good cuts. This is the first in a trio of scatter plots focused on price trends, and it transitions the viewer from physical dimension analysis into the impact of categorical features on pricing.

## Scatter Plot: Price vs Carat by Color

Similar in structure to the previous visualization, this scatter plot highlights how different color grades influence diamond pricing at similar carat weights. Higher color grades like D and E tend to cluster at higher prices with less variation, while lower grades like I and J show more price dispersion. Positioned second in the trio, this plot allows for direct comparison with the cut-based version, illustrating how color contributes to pricing patterns in nuanced ways.

## Scatter Plot: Price vs Carat by Clarity

This scatter plot visualizes how clarity grades affect the relationship between carat and price. Higher clarity diamonds like VVS1 and IF tend to fetch higher and more consistent prices, whereas lower clarity grades show more scatter and lower price ranges. As the third and final visualization in the price vs. carat series, it completes the comparative analysis across the three key categorical features, reinforcing the idea that all three contribute differently to price variability.

## Bar Chart: Average Price by Cut and Clarity

This grouped bar chart shows the average diamond price for each combination of cut and clarity. It reveals patterns such as unexpectedly high prices for SI2 diamonds in some cuts, likely due to higher carat sizes. By summarizing two attributes together, this visualization uncovers pricing interactions not visible in earlier scatter plots. It’s placed here to bridge between raw data exploration and synthesized insights, supporting more refined conclusions about diamond value.

## Best Value Recommendations Table

This table summarizes the recommended combinations of cut, color, and clarity for buyers seeking the best value. Based on patterns observed in the previous visualizations, it highlights which attributes offer the best balance between price and quality. Serving as a concluding visualization, it translates data analysis into actionable insights, making the findings more accessible to consumers or non-technical users.